METHODS AND SYSTEMS FOR PREDICTION OF A DNA PROFILE MIXTURE RATIO

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application Ser. No. 62/262,610, filed on Dec. 3, 2015 and entitled "Methods and Systems for Prediction of a DNA Profile Mixture Ratio," the entire disclosure of which is incorporated herein by reference.

GOVERNMENT FUNDING

This invention was made with Government support under Grant Numbers 2014-DN-BX-K029, awarded by the National Institute of Justice. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

The present disclosure is directed generally to methods and systems for identifying nucleic acid in a sample and, more particularly, to methods and systems for determining the ratio of contributors within a DNA mixture.

BACKGROUND

At the core of the genetic identification field, particularly in regard to forensic applications and clinical/medical research, is the challenge of DNA mixture interpretation. A DNA sample mixture can be defined as a mixture of two or more biological samples, and mastery of their interpretation can greatly impact the course of criminal investigations and/or quality of intelligence. The ability to identify the ratio of the contributors in a DNA sample may substantial improve the ability to identify the individual contributors within a mixed DNA sample.

Although historically expert systems have been in use for this problem, they often fail to meet the needs of the community, and there is continued demand by forensic communities for reliable methods of automation for mixture interpretation. The present state-of-the-art in DNA mixture interpretation includes expert systems which often have limited use, primarily focusing on improving the timeliness of analysis performed by forensic analysts. These systems capture the computational aspects of mixture analysis without taking more subjective factors into account. Further, these systems are used for simple mixtures, typically of two individuals (and thus low complexity). Although more advanced systems capable of analyzing 3-4 individual mixtures exist, these systems are both time- and cost-prohibitive.

For example, current methods to estimate the ratio of contributors in a mixed DNA sample rely on those DNA markers with the maximum number of alleles (or maximum number of alleles—1) given the number of contributors. These DNA markers have inherent variability where ratios at several loci within a sample may differ due to the size of the allele, locus base pair size, amount of degradation present, stochastic effects etc. These ratios are then used as a standard to help in the identification of the individual components or contributors in a DNA profile at those loci where a ratio cannot be determined. The calculation of these ratios is typically performed manually using a standard scientific calculator.

Accordingly, there is a need in the art for methods and systems that perform complicated DNA mixture interpretation, particularly with regard to more accurately determining the ratio of contributors within a DNA mixture.

SUMMARY OF THE INVENTION

The present disclosure is directed to methods and systems for determining the ratio of contributors within a DNA mixture. The ratio of contributors within a DNA mixture is the one of the key metrics used to separate the individual contributors during mixture deconvolution. The ratio is typically calculated using simple mathematical operations and based on known biological phenomena such as genetic dosage. This method, although effective, is limited by the capacity of human computation and fails to utilize much of the information contained within the profile.

Accordingly, the methods and systems described herein combine statistical and biological approaches which are made feasible through a processor. According to an embodiment, the system includes a combinatorial algorithm to enumerate all potential DNA mixture scenarios within a single DNA marker. The system further includes an outlier removal algorithm, and a clustering algorithm to identify the most similar ratios among DNA markers.

According to one aspect is a method for determining a ratio of the proportion of DNA from each contributor within a mixed DNA sample, comprising the steps of: (i) characterizing a parameter of the DNA mixture; (ii) characterizing a plurality of markers within the DNA mixture; (iii) identifying which of the plurality of markers exhibits a maximum number of alleles, wherein at least one of the plurality of markers is identified; (iv) enumerating, based on the identification, all possible scenarios for contributors to the DNA mixture; (v) determining a mixture ratio for each enumerated scenario, wherein every allele found in a given marker must be represented in the scenario; (vi) identifying all possible clusters for the determined mixture ratios, wherein a cluster is a group of ratios comprising just one ratio from the at least one identified marker; (vii) removing any statistical outliers from each of the identified clusters; (viii) identifying candidate clusters, wherein a cluster is identified as a candidate if the variance of the distance from each mixture ratio to the cluster's centroid is below a certain user-specified threshold; and (ix) comparing all of the candidate clusters to all the of the mixture ratios, wherein the candidate ratio with the highest number of markers containing at least one similar ratio at each marker is identified as the DNA profile mixture ratio.

According to an embodiment, the method further includes the step of characterizing a parameter of the DNA mixture.

According to an embodiment, the method further includes the step of characterizing the plurality of markers within the DNA mixture.

According to an embodiment, the method further includes the step of preparing the sample for analysis.

According to a second aspect is a system configured to characterize a ratio of contributors to a DNA mixture within a sample. The system includes: a sample preparation module configured to generate initial data about the DNA mixture within the sample; a processor comprising a ratio of contributors determination module, the ratio of contributors determination module configured to: (i) receive the generated initial data; (ii) analyze the generated initial data to determine the ratio of contributors to the DNA mixture within the sample; and an output device configured to receive the determined ratio of contributors from the pro-